Cluster Analysis of Air Quality Data for CCOS Study Domain

Ahmet Palazoglu, PI
University of California, Davis, CA

Presented by Scott Beaver
Bay Area Air Quality Management District

CCOS TC Meeting
4 November 2008
EPA Building, Sacramento, CA

3 Purposes of Presentation

1. Summary of clustering work for CCOS domain
   - 6 subdomains: Bay Area, San Joaquin Valley (North, Central, and South), Sacramento Valley & Mountain Counties
   - 1996-2004 “extended” ozone seasons (1 May -- 31 Oct)
   - Separate clusterings of wind & O_3 measurements
2. Domain-wide analysis
   - Qualitative agreement of results between subdomains
     - Similar synoptic influences for all subdomains
     - Example episodic scenario affecting entire domain
   - Quantitative analysis not feasible
     - Data limitations
3. Extensions of methods
   - Practical applications for modeling (simulation) efforts
   - Proofs of concept demonstrated for SFBA

Work Plan Tasks

1. Wind field clustering for 6 CCOS subdomains
   - ✓ Collect data
     - ✓ Non-contracted but necessary quality assurance delayed overall contract progress
   - ✓ Perform 6 independent clusterings: SFBA, N/C/S SJV, SV, MC
   - ✓ Interpret clustering results
2. O_3 clustering for 6 CCOS subdomains
   - ✓ Collect data
   - ✓ Test alternative algorithms for SFBA
   - ✓ Perform clusterings for other 5 subregions: N/C/S SJV, SV, MC
   - ✓ Interpret clustering results
3. Sequencing of dynamic cluster patterns
   - ✓ Attempt to relate wind and O_3 clusterings
   - ✓ Identify recurring upper-atmospheric transitions
   - ✓ Consider dynamics at other time scales
4. Domain-wide synopsis of wind field clusterings
   - ❑ Data limitations precluded quantification of these results

Lessons Learned

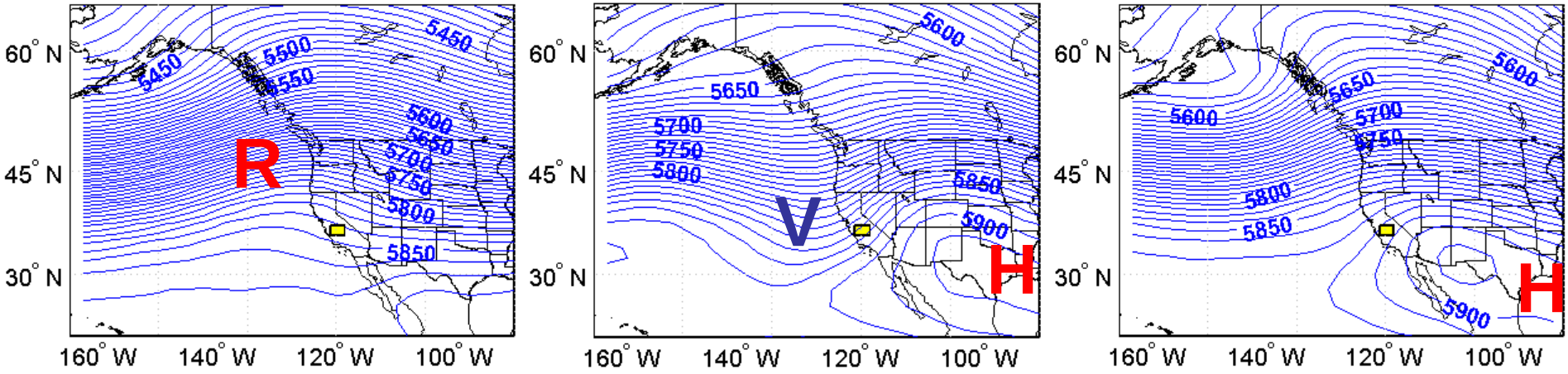
- Wind field clustering & sequencing contribute significant information
  - Provide physical insight for CCOS domain ozone episodes
  - Can provide increased representativeness and confidence in modeling (simulation) efforts
  - Methods useful for winter PM analysis (ongoing BAAQMD contract)
- Multi-scale nature of CCOS domain ozone variability
  - Mesoscale flow features further refine synoptically oriented clusters
  - Sources of inter-annual variability are not easily resolved
- Complexity of meteorology varies by basin
  - SFBA least complex
    - Slight amount of marine ventilation inhibits ozone buildup
  - Sacramento Valley most complex
    - Bi-directional flows along valley major axis
    - Small changes in marine ventilation (not affecting other basins) may strongly impact ozone levels
  - SJV has considerable spatial variability
    - Thermal flows stronger deeper south into SJV; marine influences stronger toward north
    - Fresno Eddy generates very complex but localized flow patterns
- Novel quality assurance methods were developed and applied to significantly enhance the results
- O_3 measurements clustering is not very useful
- Data limitations unknown at project outset precluded quantitative domain-wide analysis
  - Study could have been designed to focus on domain-wide patterns at expense of lost detail for individual subdomains

Consistent clusters for 5 subdomains

Number of days (and % NAAQS 8-hr O_3 exceedances) for each cluster, by subdomain.

	SFBA	N-SJV	C-SJV	S-SJV	SV
R	86 (13%)	264 (19%)	370 (60%)	404 (61%)	302 (35%)
H	353 (13%)	179 (42%)	229 (89%)	429 (58%)	397 (28%)
H/V	---	212 (25%)	184 (64%)	249 (63%)	203 (17%)
V or V1	309 (0%)	299 (12%)	299 (42%)	193 (38%)	228 (3%)
V/I or V2	341 (0%)	108 (6%)	396 (32%)	335 (22%)	171 (12%)
V/R	---	---	---	---	200 (3%)

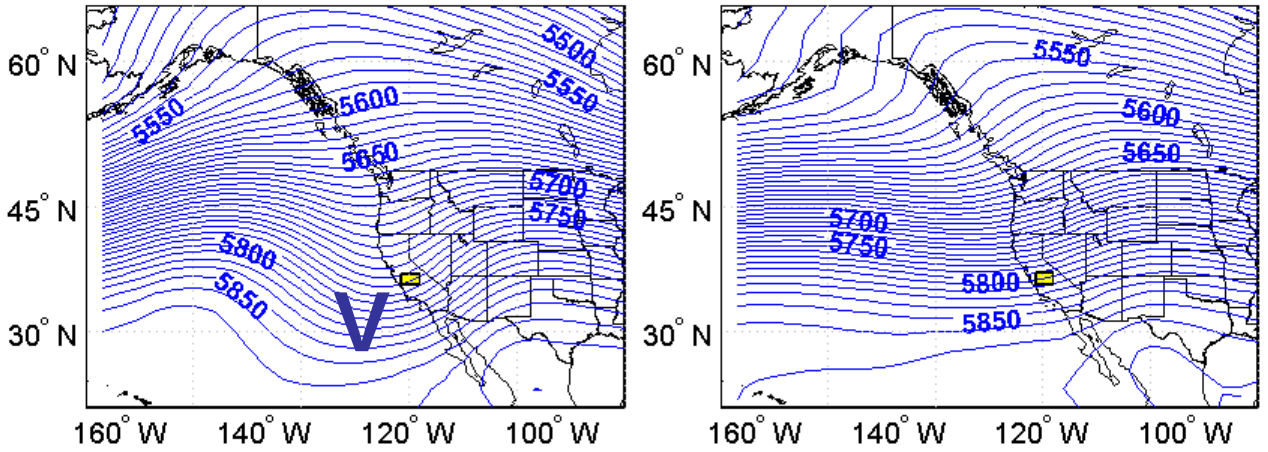
“Static” cluster patterns (500-hPa)



R H/V H

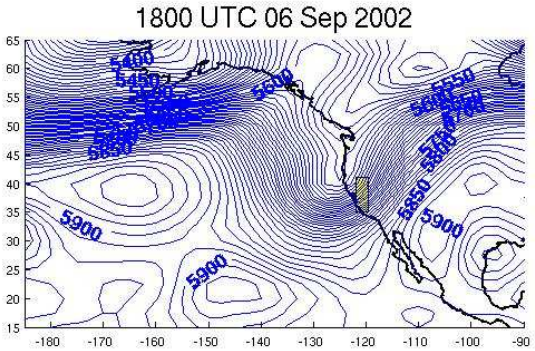
Similar synoptic influences for all subdomains, but surface flows vary considerably.

V or V1 V/I or V2

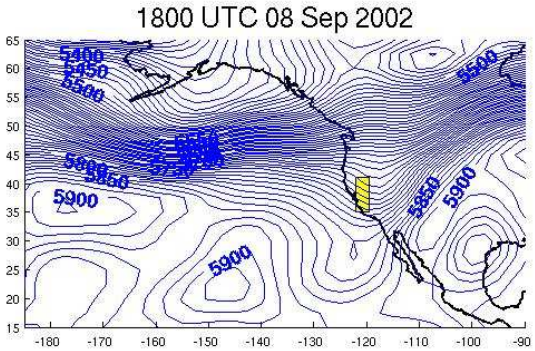


“Dynamic” cluster sequences (500-hPa)

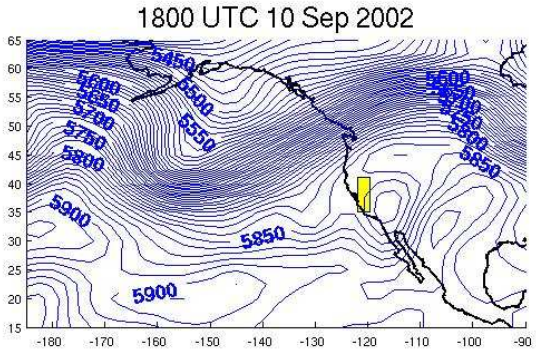
Example: V→R→H captures eastward sweeping Rossby wave



Trough along coast buffers CCOS domain from effects of offshore high pressure systems

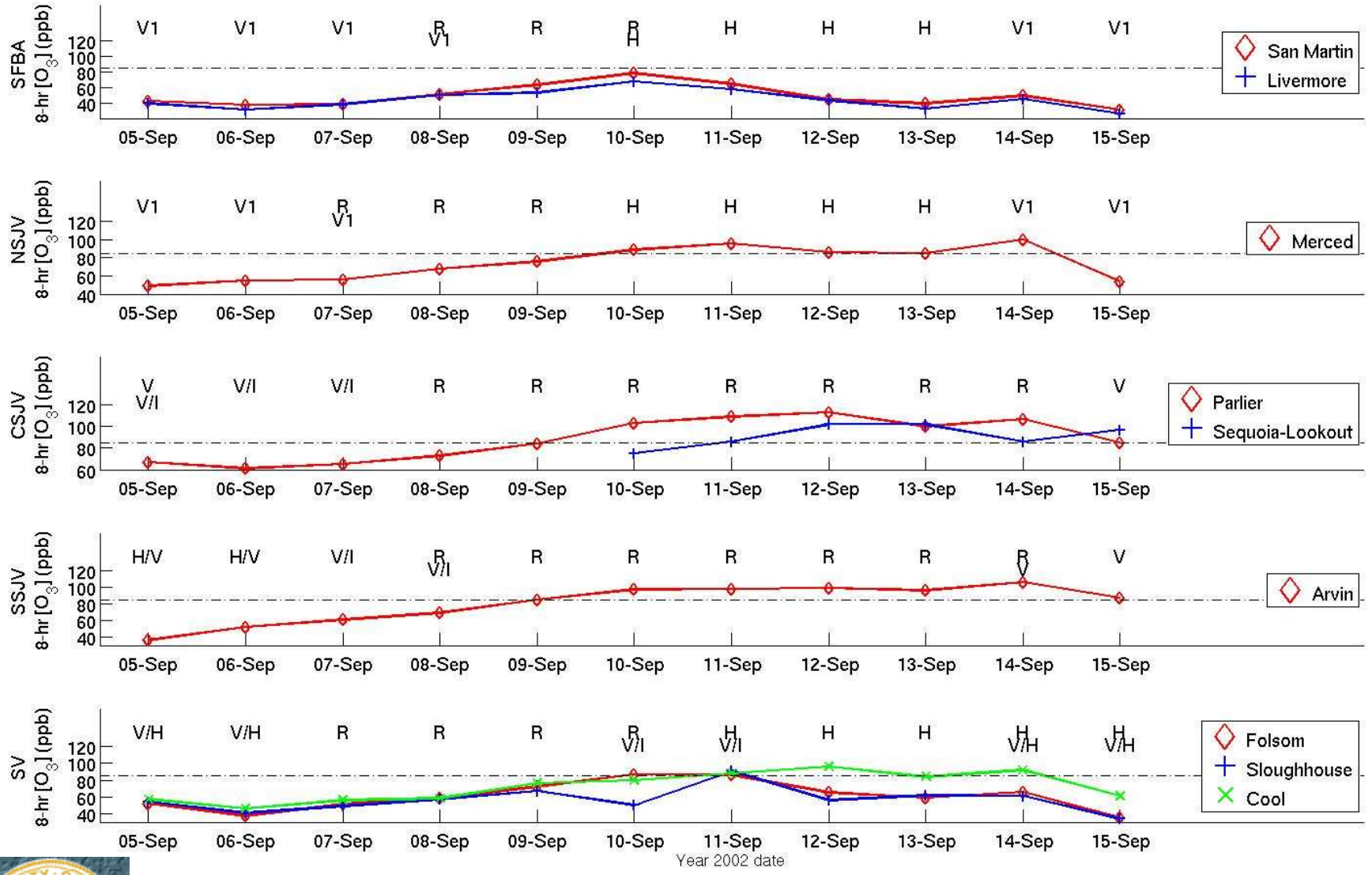


Trough dissipates as offshore high pressure advances eastward



Ridge forms along coast from high pressure of offshore origin

“Dynamic” cluster sequences (surface)



Domain-wide Hypothetical Example

	SFBA label	NSJV label	CSJV label	SSJV label	SV label	All labels present?
day 1	V		V		V	
day 2	V		V		V	
day 3	R					
day 4	R	R	R	H/V	R	YES
day 5	R	R		R	R	
day 6	R	R		R	R	
day 7	R	R		R	R	
day 8	H	R	R	R	H	YES
day 9	H	H	H	R	H	YES
day 10		H	H	H	H	
day 11		H	H	H	H	
day 12		H	H	H	H	
day 13	H	H	H	H		

Tradeoff between spatial and temporal resolution for each subdomain.
Only 452 of 1656 days in cluster analysis have labels for all subdomains!
Lowered spatial resolution could increase domain-wide sample sizes.

Practical Extensions of Clustering

- Winter PM season clustering & sequencing
  - BAAQMD contract with UC Davis provides proof of concept
  - Initial results useful for modeling efforts
- Meteorological & Air Quality Model (AQM) validation
  - Classify simulated winds among known patterns
  - Determine if simulated data (classification) are labeled consistently with observations (clustering)
    - Prevailing conditions
    - Upper-atmospheric transitions
  - Expect better AQM performance when met. modeled accurately
  - Determine if seasonal met model can explain air quality variability
    - Proof of concept for MM5 winter 2000-01 simulation at BAAQMD
    - Compare MM5 performance to different met. models (WRF, CALMET)
- Selection of representative conditions for future simulations

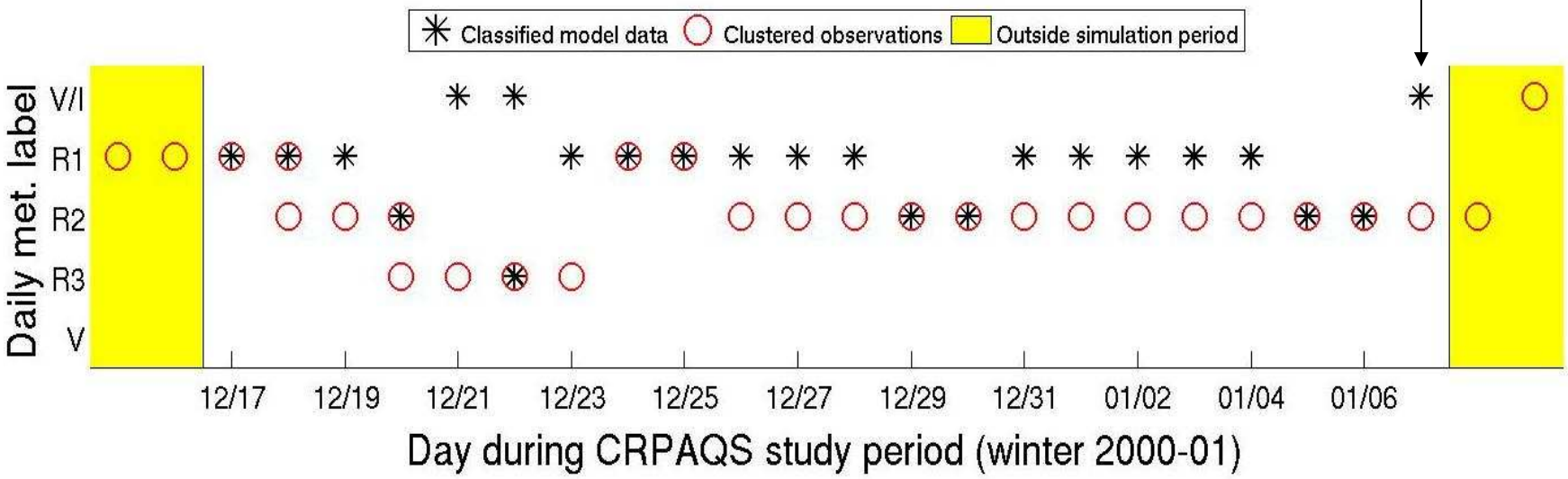
- [blank slide]

MM5 validation: SFBA example

First half of CRPAQS episode: R1→R2→R3→R1 realistically simulated.

Second half of CRPAQS episode: persistent R2 inaccurately simulated as R1.

Simulated ventilation arrives 1-2 days early.



The 12/17/2000-1/7/2001 CRPAQS study period is simulated using MM5. This period was included in a previous UC Davis cluster analysis.

CMAQ performance evaluation

First half: adequate performance

Second half: PM levels underestimated.

Last day: PM levels decrease too early for eastern sites